Multiscale Data Assimilation

Dr. Pierre F.J. Lermusiaux
Department of Mechanical Engineering, Center for Ocean Science and Engineering,
Massachusetts Institute of Technology; 5-207B; 77 Mass. Avenue; Cambridge, MA 02139-4307
phone: (617) 324-5172 fax: (617) 324-3451 email: pierrel@mit.edu

Grant Number: N00014-13-1-0518 FY14 report
04/01/2013 - 09/30/2015
http://mseas.mit.edu/

LONG-TERM GOALS

This research is concerned with next-generation multiscale data assimilation, with a focus on shelfbreak regions, including non-hydrostatic effects. Our long-term goals are to:
- Develop and utilize GMM-DO data assimilation schemes for rigorous multiscale inferences, where observations provide information on varied spatial and temporal scales
- Develop and utilize test cases and simulation experiments that allow the evaluation of such schemes in multiscale dynamics conditions, including non-hydrostatic processes in shelfbreak regions.

OBJECTIVES

The specific objectives are to:
- Further develop, illustrate and determine the capabilities of the GMM-DO filter for multiscale data assimilation.
- Develop and utilize test cases and simulation experiments for the evaluation of data assimilation schemes in multiscale dynamics conditions.
- Study the multiscale properties of probability density functions predicted by GMM-DO, including multiple scales in time and multiple scales in space.
- Based on these properties, develop multi-resolution measurement operators and possibly multi-resolution GMM-DO filters and smoothers.
- Strengthen collaborations, transferring our test cases for multiscale data assimilation and our approaches to NRL. Utilize and leverage the MIT Naval Officer education program.

APPROACH

While traditionally grounded in linear theory and the Gaussian approximation, one recent research thrust for data assimilation has been the development of efficient assimilation methods that respect nonlinear dynamics and capture non-Gaussian features. Most such methods are either challenging to employ with large realistic systems or still based on heuristic hypotheses and ad hoc approximations. Our unique motivation here is to allow for realistic multiscale dynamics while rigorously utilizing the governing dynamical equations with information theory and learning theory for efficient Bayesian inference. To do so, we employ the recent results of the MSEAS-group in such equation-based non-

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **30 SEP 2014**	2. REPORT TYPE	3. DATES COVERED **00-00-2014 to 00-00-2014**

4. TITLE AND SUBTITLE **Multiscale Data Assimilation**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Massachusetts Institute of Technology,Department of Mechanical Engineering,Cambridge,MA,02139**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Gaussian data assimilation, combining the stochastic Dynamically Orthogonal (DO) field equations with semi-parametric Gaussian Mixture Models (GMMs). The challenge of our research is to allow for truly multiscale inferences, where observations and models provide information on varied spatial and temporal scales.

For our multiscale data assimilation, two complementary approaches are investigated. The first approach is direct multiscale filtering and smoothing which starts with and extends our new GMM-DO nonlinear filter. The second approach is based on arguments of scale-decomposition, and even scale separation if such separations can be justified. Important feedbacks are multiscale adaptive sampling and adaptive modeling. For researching and developing next-generation multiscale data assimilation ideas, we utilize our MIT-MSEAS modular, flexible framework in Python and Matlab which has been developed specifically for such incubation purposes. Observation system simulation experiments are developed and utilized, focusing on tidal to regional ocean processes occurring at shelfbreaks.

WORK COMPLETED (FY14)

GMM-DO Codes: Numerical Improvements, GMM Fits and DO Closure. Several numerical studies for our DO codes were completed. Our group also completed a detailed review of our GMM-DO codes, identifying a set of numerical studies: e.g. stability for DO numerics (CFL), proper upwind advection for DO modes, DO normalization and re-orthonormalization, and single multivariate DO coefficients versus multiple univariate DO coefficients, DO numerical closure. Several of these studies can be carried out in the last year of this project. For the present year (e.g. Lin, 2015), we improved the boundary condition discretization of our FV non-hydrostatic code (which is used for GMM-DO studies) to second order accurate and verified convergence rates. We also improved the set-up and efficiency of our LU linear solve. We also coded DO terms for anisotropic diffusion and developed a modified incremental projection method with rotational correction for this purpose. For the fitting of DO forecasts to GMMs, we explored a set of different fitting strategies, including subspace clustering, parsimonious GMM and sparse GMM by l1 penalization. We also evaluated the convergence of the DO pdf predictions in terms of the dimension of the stochastic subspace employed, and outlined strategies for efficient DO closures.

Test Cases for Multiscale Data Assimilation: Several test cases involving varied dynamics were developed. Two of them include non-hydrostatic flows behind a seamount (Fig. 1) and non-hydrostatic bottom gravity currents (Fig. 2). In the seamount test case, flows with varying Reynolds number were studied. The resulting different parameter regimes highlight different multiscale physics at the seamount including vortex generation, lee waves and unstable flows (to name a few). These different flow regimes are currently being used to test our GMM-DO assimilation schemes. In the bottom gravity current test case, the multiscale phenomena in a flow of salty water from a plateau over a linear slope are simulated. Fig. 2, which contains contour plots of salinity at different time instances, shows the deterministic simulation setups and results. The domain geometry is formed by a horizontal plateau followed by a 2.5-degree linear slope. Open boundary conditions are applied to both the inlet and the outlet. A no-slip boundary condition is applied at the bottom, a free-slip at the top. Initially (Fig. 2(a)), there is no flow and salinity anomaly is set over a portion of the plateau. The dynamics of this density-driven flow mainly depends on the Grashof number, which is Gr=759 for the simulation shown in Fig. 2. In the beginning (Fig. 2(b) and (c)), the salty water goes down the slope mainly as a simple shear flow led by a gradually developing head structure. Later (Fig. 2(d)), the Kelvin-Helmholtz instability is triggered and billows are formed from the shear flow. Finally (Fig. 2(e)), the billows begin to interact with each other and with the head, and finer-scale structures are formed. Therefore, this is a good test

case to study the multiscale data assimilation capabilities of our GMM-DO filter. We also performed stochastic simulations with our DO equations and data assimilation with our GMM-DO filter for this bottom gravity current (Fig. 3).

Skill of Multiscale Ocean Probability Forecasts: Prior to use in realistic multiscale data assimilation, ocean probability forecasts need to be evaluated. This was completed (Lermusiaux et al, 2015) using multiscale observations collected during the two-month QPE IOP09 real-time experiment off the coast of Taiwan during Aug-Sep 2009. The multiscale ESSE ensemble forecasts were compared to the measured errors between the central forecasts and seasoar data (which contains strong nonlinear internal waves and tides), and also between the central forecasts and the objectively analyzed seasoar data (so as to filter the faster and shorter scales, e.g. due to these strong internal waves and tides). The variability in the pdfs were illustrated and discussed, including effects of Typhoon Morakot and internal tides. The ignorance score and Kullback-Leibler divergence were employed to measure the skill of the multiscale pdf forecasts.

Multiscale Smoothing with the GMM-DO Smoother: We further developed our GMM-DO smoother, building on concepts from our GMM-DO filter. The smoother uses the DO equations for uncertainty prediction and the GMM-DO scheme for filtering. Smoothing is performed using a state augmentation procedure in which the past and the present states are first appended to form the prior distribution of a larger state vector. Observations are then assimilated by efficiently carrying out Bayes law in the reduced DO subspace of the augmented vector, using our GMM-DO filter. The smoothed distribution is then read off from the posterior of the augmented state vector. We implemented this new smoother and tested it using a 2D-in-space stochastic flow exiting an idealized ocean strait.

RESULTS (FY14)

GMM-DO Codes: Numerical Improvements, GMM Fits and DO Closure. We found that the improvements made to our DO code were as expected and allowed more accurate simulations of pdfs for multiscale ocean flows. The study of GMM fits indicated that our ideas of fitting the dominant portion of the subspace with full GMMs while fitting the remainder with a single (but broad) Gaussian would likely be much better than other techniques used in the literature for other (non-fluid dynamics) problems. We also confirmed that a DO closure is needed for (too) small subspace.

Test Cases for Multiscale Data Assimilation: In the flow past a seamount test (Fig. 1), different flow regimes, with their different physical scales, are explored using various values of the Reynold's number (Re). In (Fig. 1a) a laminar flow develops for Re=1. The steady state shown in Fig. 1a develops quickly (fully developed by nondimensional time t=10). In Fig. 1b, a recirculation gyre forms in the lee of the mountain for a moderate flow (Re=100). The steady state takes three times longer to develop. A stronger flow (Re=10^6) is shown in Fig. 1c. The flow in the wake of the seamount quickly becomes unsteady. Smaller scale vortices and lee waves are shed by the seamount. In the deterministic bottom gravity current test case (Fig. 2), multiscale physics develops along with time, from simple shear flow to complicated billows formed by Kelvin-Helmholtz instabilities. Results of stochastic bottom gravity current runs with data assimilation performed once a minute from T=30min are demonstrated in Fig. 3. Fig. 3(a) and 3(b) show the results before data assimilation at T=30min and T=45min, respectively. The first row in Fig. 3(a) and Fig. 3(b) show the mean u velocity field (left), the mean salinity field (middle) and the time evolution of the mode variances. In the additional five rows of Fig. 3(a), the first five DO modes are illustrated through the modes of u velocity (left column), the modes of salinity (middle column) and the p.d.f.s of stochastic coefficients (right column). We can

see how the multiscale flow structures are reasonably captured by the DO modes and how the non-Gaussian (multi-modal and skewed) statistics are captured by the DO coefficients. The right plot in Fig. 3(b) shows how the growth of mode variances due to chaotic dynamics is controlled by the GMM-DO data assimilation and levels off.

Skill of Multiscale Ocean Probability Forecasts: Using multiscale observations collected during the QPE real-time experiment, the multiscale ESSE ensemble forecasts were compared to the measured errors between the central forecasts and seasoar data, and also between the central forecasts and the objectively analyzed seasoar data (Lermusiaux et al., 2015). RMS statistics showed a good agreement between forecast and measured errors. In doing so, a real-time numerical bias was removed by improving/correcting the deterministic MSEAS primitive-equation code with a free surface. Pdfs of the forecast errors were shown to capture and evolve non-Gaussian and multiscale statistics, corresponding to the pdfs of larger-scale flows, mesoscale features including meanders and eddies, internal tides, and barotropic tides. Comparing the Kullback-Leibler divergences for the forecast error pdfs with a climatological pdf distribution showed that our forecast pdfs improved the climatology pdf by 50 to 100%. The ignorance score showed 25 to 50% improvement in the forecast pdfs over the climatology pdf. We also found that adding a stochastic tidal forcing strongly affected the forecasts of velocity pdfs on the shelf. Our reanalysis with improved numerics and parameters removed deterministic biases and improved pdf comparisons.

Multiscale Smoothing with the GMM-DO Smoother: Our identical twin experiments with a dynamic 2D-in-space flow exiting a strait showed that qualitatively, our new GMM-DO smoother was very effective at estimating 2D-in-space currents backward in time from a limited number of flow measurements. The smoother estimates were found to resemble the true solutions very well. Quantitatively, we also found that the smoothed field had RMSEs that decayed quickly with the number of observations assimilated. The smoothed field converged towards the true field and critically, its posterior statistics also converged towards the statistics of the true errors (smoothed field minus true field).

IMPACT/APPLICATIONS

New multiscale non-Gaussian data assimilation methods are critical for major improvements of ocean forecasting systems and directly relevant to naval interests. Other major impacts are expected on scientific, naval and societal activities that involve multiscale ocean processes; coupled physics, acoustics, ecosystem or sea-ice dynamics; weather and atmospheric dynamics, and climate dynamics.

TRANSITIONS AND COLLABORATIONS

We plan to collaborate with NRL and colleagues to develop, demonstrate and transfer ideas and approaches for multiscale data assimilation. During the last year of the project, our results can be tested in more realistic ocean simulations and at-sea experiments of opportunity. Possibilities include the Mid-Atlantic Bight and Shelfbreak Front region, the Chinese-Taiwanese Seas, the Philippine Seas, the Massachusetts Bay/New England shelf region, and the Monterey Bay and California Current System region. For the sea exercises, possibilities include NATO exercises with the NURC. To provide efficient education, we plan to leverage the MIT Naval officer education program so as to continue to attract METOC officers and practitioners, either for focused shorter visits (e.g. in the summer), or for Master’s or PhD degrees.

RELATED PROJECTS

Related research projects include: N00014-13-1-0514 (B. Powell), N00014-13-1-0520 (B. Cornuelle) N0001413-WX21102/RX20289 (E. Coelho and K. Heaney). Our project on Active Transfer Learning for Ocean Modeling (N00014-11-1-0337) also benefits from the test case we develop for the present study.

STUDENT SUPPORTED: This small project supported the equivalent of one graduate student one third of the time. One METOC officer, Jen Landry (LCDR USN) who was directly supported by the Navy, also learned from the project and completed a SM thesis. A summer visiting student from India, A. Gupta, also contributed to the project.

PUBLICATIONS

Lermusiaux P.F.J., P.J. Haley Jr. and G.G. Gawarkiewicz, 2015. *Evaluation of Multiscale Ocean Probabilistic Forecasts: Quantifying, Predicting and Exploiting Uncertainty*. To be submitted to the Journal of Ocean Dynamics.

THESES

Landry, J.J., 2014. *Coastal Ocean Variability off the Coast of Taiwan in Response to Typhoon Morakot: River Forcing, Atmospheric Forcing and Cold Dome Dynamics*. SM Thesis, MIT-WHOI Joint Program, September 2014

Lin J., 2015. *Bayesian Learning for Multiscale Ocean Flows*, SM Thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, February 2015. Expected.

Other publications are in preparation. Additional presentations and other publications are available from http://mseas.mit.edu/. Other specific figures are available upon request.

FIGURE

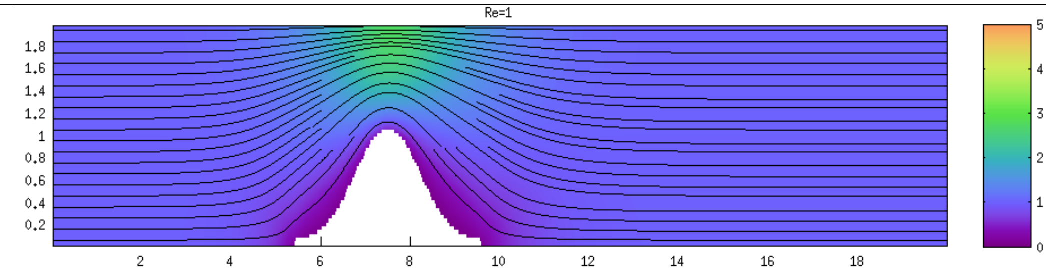
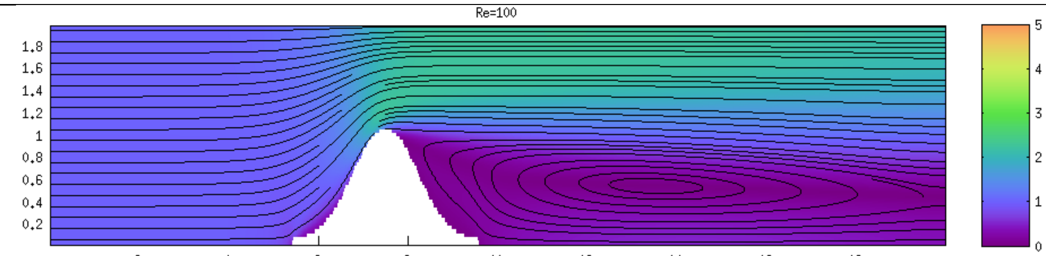
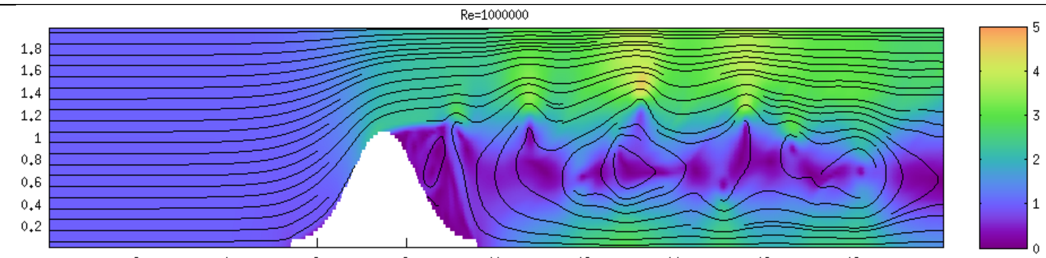
(a) Re=1	
(b) Re=100	
(c) Re=10^6	

Figure 1: Flow past a seamount. Different flow regimes are explored using various Reynold's numbers (Re). (a) Re=1, laminar flow. Steady state (shown) is reached quickly (nondimensional time t=10). (b) Re=100 a recirculation gyre forms in the lee of the mountain. Steady state (shown) is reached more slowly (by t=30). (c) Re=10^6, unsteady flow. Vortices and lee waves are shed by the seamount.

(a) T=0h

(b) T=1h

(c) T=2h

(d) T=3h

(e) T=4h

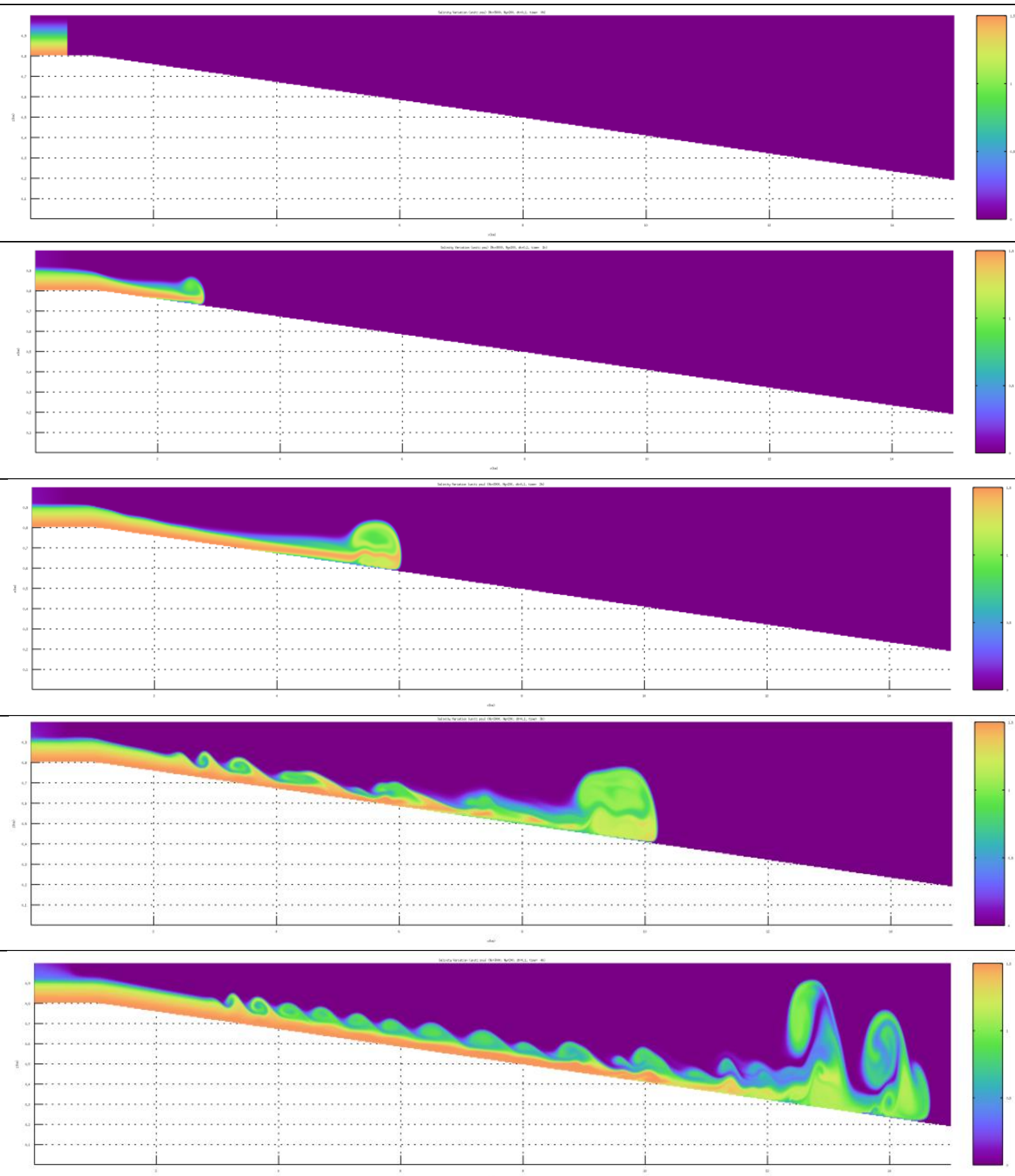


Figure 2: Salinity contour plots of a deterministic bottom gravity current simulation. The domain shown in the plots is 15km long and 1km high. The 0.8km-high plateau is followed by a 2.5-degree linear slope. The Grashof number is 759. (a) T=0h. Initial salinity anomaly over the plateau. (b) T=1h. Simple shear flow led by small head. (c) T=2h. Further developing shear flow and head structure with finer scales. (d) T=3h. Kelvin-Helmholtz stability triggered. (e) T=4h. Further developing multiscale flow structures.

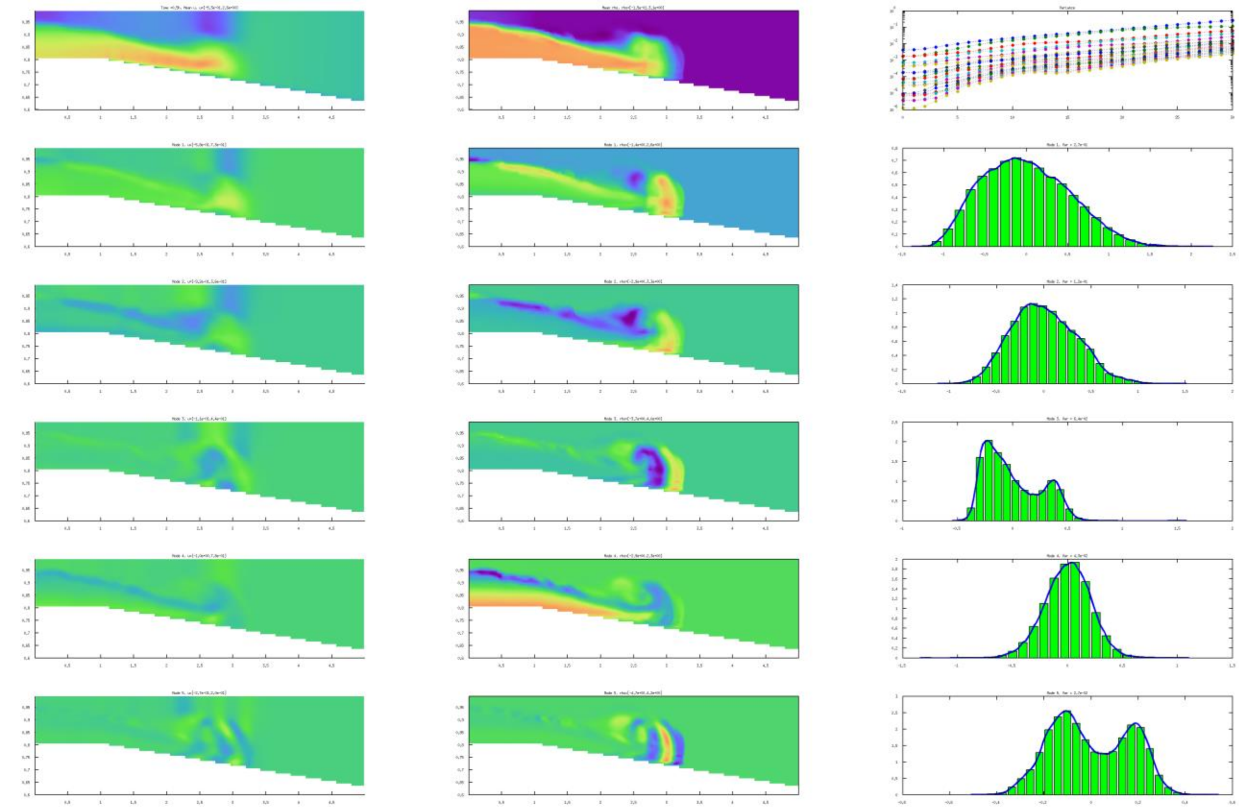
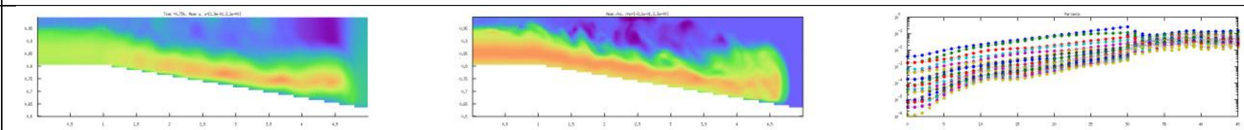
(a) T=30 min	
(b) T=45 min	

Figure 3: Simulation results of a stochastic bottom gravity current with data assimilation. The domain shown in the plots is 5km long and 1km high. The 0.8km-high plateau is followed by a 2.5-degree linear slope. The Grashof number is 1518. (a) T=30min. The first row shows the u velocity mean (left), the salinity mean (middle) and the time evolution of mode variances (right). The next five rows show the u velocity modes (left column), the salinity modes (middle column) and p.d.f.s of stochastic coefficients (right column). (b) T=45min. The three plots are the u velocity mean (left), the salinity mean (middle) and the time evolution of mode variances (right).